IMPROVED SET ESTIMATORS FOR A MULTIVARIATE NORMAL MEAN

Short Title: Improved Set Estimators

BY JIUNN TZON HWANG[1] AND GEORGE CASELLA[2]

Cornell University

September, 1982

[1]Research supported by National Science Foundation Grant No. MCS80-03568.

[2]Research supported by National Science Foundation Grant No. MCS81-02541.

AMS 1980 subject classifications. Primary 62F25; 62J07; secondary 62C20.

Key words and phrases. Confidence sets; Stein estimation; multivariate normal density.

Abstract

The knowledge of improving upon the usual point estimator (the maximum likelihood estimator) of the mean of a p-variate normal distribution ($p \geq 3$) is, by now, almost complete. The theoretical development of improving upon the usual confidence set has been slow, partly due to the difficulty of proving domination in coverage probability.

This paper provides an alternative to the proof of Hwang and Casella (1982) for the domination of the sphere recentered at the positive-part James-Stein estimator over the usual confidence set. The alternative proof also yields stronger results. The proof is based on a fairly simple and general formula which applied to some other spherically symmetric distributions. A necessary condition for domination is also obtained.

§1. Introduction.

Assume that X is a p-variate normal random variable with mean θ and identity covariance matrix. Stein (1956) proved that the usual point estimator of $\theta, \delta^0(X) = X$, can be improved under the sum of squared error loss if $p \geq 3$. Since then a tremendous amount of work has been done by many statisticians to improve upon δ^0. In particular, it was shown that the positive-part version of the James and Stein (1961) estimator,

$$\delta^a = \left(1 - \frac{a}{|X|^2}\right)^+ X \quad , \quad 0 < a < 2(p-2) \tag{1.1}$$

dominates δ^0. (See Efron and Morris 1973.) In equation (1.1), "+" denotes the positive part and $|Y|$ denote the Euclidean norm of a vector Y.

In estimating θ, we need, not only a point estimator, but also a confidence set as well. The classical confidence set C_X^0 associated with δ^0 is a sphere centered at δ^0, i.e.

$$C_X^0 = \{\theta : |X-\theta| \leq c\} \quad .$$

The radius c of C_X^0 is usually chosen so that

$$P(\chi_p^2 \leq c^2) = 1-\alpha \quad ,$$

where χ_p^2 is a chi-squared random variable with p degrees of freedom. This implies that the coverage probability of C_X^0

$$P_\theta(\theta \in C_X^0) = 1-\alpha \text{ , for all } \theta \text{ .}$$

A confidence set C_X^1 will be said to dominate (or to uniformly improve upon) C_X^2 if, for all θ and X,

(i) $P_\theta(\theta \in C_X^1) \geq P_\theta(\theta \in C_X^2)$

and

(ii) Volume of $C_X^1 \leq$ volume of C_X^2

with strict inequality holding in either (i) or (ii) for a set of θ or X, with positive Lebesgue measure.

Associated with an improved point estimator, one usually needs a confidence set. Since the improved point estimator is closer than δ^0 to θ on the average, it is reasonable to hope that the associated confidence set will also improve upon C_X^0 uniformly.

Stein (1962) gave some strong heuristic arguments to indicate that C_X^0 is dominated, for large p, by the confidence set recentered at δ^a , i.e. the confidence set

$$C_{\delta^a} = \{\theta : |\theta-\delta^a(X)| \leq c\} .$$

Brown (1966) and Joshi (1967) independently prove that C_X^0 can be improved uniformly for $p \geq 3$. Their proofs are existential, however. Joshi (1969) later proved that C_X^0 can not be improved if $p = 1$ or 2.

Attempts to construct specific confidence sets dominating C_X^0 continue. By using a version of Bayes sets, Faith (1976) derived alternative confidence sets. He proved that his confidence sets improve upon C_X^0 for certain regions of $|\theta|$ and $|X|$. His numerical study indicates that these confidence sets improve upon C_X^0 uniformly. It is not clear, however, what point estimator is associated with his confidence set. Berger (1980) developed confidence sets, associated with his admissible minimax generalized Bayes point estimator, through the consideration of a posterior mean and a posterior covariance matrix. He also gave convincing analytical and numerical evidence that his confidence sets improve upon C_X^0 uniformly. These confidence sets, although performing very satisfactorily, are diffcult to calculate.

Simpler confidence spheres centered at δ^a were considered in Casella (1980), Hwang and Casella (1982), and Casella and Hwang (1982). Casella (1980) dealt with $C_\delta a$, a sphere of the same radius c. The confidence sphere in Casella and Hwang (1982) has a radius, depending on $|X|$, which is always smaller than c. They presented numerical evidence to show the domination of their confidence sets over C_X^0.

The first analtical constructive results for the domination of C_X^0 were obtained in Hwang and Casella (1982). It was proved that $C_\delta a$ has higher coverage probability than C_X^0 for all θ provided $p \geq 4$ and $0 < a \leq a_0$. The constant a_0 is approximately $0.8(p-3)$, and is given in Table 1. Since the volume of $C_\delta a$ is the same as that of C_X^0, it follows that $C_\delta a$ dominates C_X^0.

The proof in Hwang and Casella (1982) is based on a formula for $\frac{\partial}{\partial|\theta|} P_\theta(\theta \varepsilon C_\delta a)$, which is established by some lengthy calculations and applications of Dirac delta functions. Furthermore, many questions related to the domination of C_X^0 remain unsolved. In particular,

(1) no analytical constructive results are obtained for $p = 3$;

(2) numerical results in Casella (1980) indicate that the theorem should remain true even for a larger a_0. Moreover, for $a < p-2$, the corresponding $C_\delta a$ can be improved uniformly by increasing a. Therefore, it should be possible to strengthen their theorem by enlarging a_0.

In this paper, an alternative proof of the domination of $C_\delta a$ over C_X^0 is given. The theorem established here is also stronger than the results in Hwang and Casella (1982) in that it solves, to some extent, both problems (1) and (2).

The proof is based on a formula for $\frac{\partial}{\partial a} P_\theta\ (\theta \varepsilon C_\delta a)$, rather than $\frac{\partial}{\partial|\theta|} P_\theta\ (\theta \varepsilon C_\delta a)$. It is then shown that the derivative is positive for a certain range of a $(0 < a \leq a^*)$, which implies that $P_\theta(\theta \varepsilon C_\delta +)$ is increasing in a. Note that when a=0, C_{δ^a} reduces to C_X^0. Consequently, for $0 < a \leq a^*$, $P_\theta(\theta \varepsilon C_\delta a) > P_\theta(\theta \varepsilon C_X^0)$.

The derivative formula is quite simple and is fairly general in that it applies to spherically symmetric distributions other than the normal. Therefore the technique should prove to be useful in future studies.

Section III establishes an asymtotic formula (as $|\theta| \to \infty$) for the coverage probability of $C_\delta a$. This formula implies that a necessary condition for the domination of $C_\delta a$ over C_X^0 is $0 < a \leq 2(p-2)$.

§2. A Derivative Formula and a Sufficient Condition.

In Lemma 1 below, we derive a formula for $\frac{\partial}{\partial a} P_\theta(\theta \in C_\delta a)$. This formula is used to establish a sufficient condition for domination of $C_\delta a$ over C_X^0 . Let $h_\theta(a)$ denote the coverage probability $P_\theta(\theta \in C_\delta a) = P_\theta(|\theta-\delta^a(X)| \leq c)$. Then

$$h_\theta(a) = \int_{|\theta-\delta^a(X)| \leq c} f(X-\theta)dX$$

where $f(X-\theta)$ is the p-variate normal density with mean θ and identity covariance matrix.

Before we prove Lemma 1, we first derive an expression for $h_\theta(a)$ by a spherical transformation. Specifically, let $r = |X|$ and let β be the angle between X and θ . The inequality

$$(2.1) \qquad |\theta - \delta^a(X)| \leq c$$

is then equivalent to

$$(2.2) \qquad r^2u^2(r) - 2ru(r)|\theta|\cos\beta + |\theta|^2 \leq c^2$$

where $u(r) = (1 - \frac{a}{r^2})^+$. Rewriting (2.2) and a little algebra show that the set of X satisfying (2.1) equals the region $\{x: r_- \leq r \leq r_+ \text{ and } 0 \leq \beta \leq \beta_0\}$. For $|\theta| \leq c$, $\beta_0 = \pi$, $r_- = 0$, and $r_+ = r_+(a,\theta,\beta)$ satisfies

$$(2.3) \qquad r_+u(r_+) = |\theta|\cos\beta + \sqrt{c^2-|\theta|^2\sin^2\beta} \; ;$$

for $|\theta| > c$, $\beta_0 = \sin^{-1}(c/|\theta|)$ and $r_\pm = r_\pm(\theta,a,\beta)$ are solutions to

$$r_{\pm}u(r_{\pm}) = |\theta|\cos\beta \pm \sqrt{c^2-|\theta|^2\sin^2\beta} \overset{\text{def'n.}}{=} r_{\pm}^0 ,$$

i.e.,

$$r_{\pm}(\theta,a,\beta) = [r_{\pm}^0 \pm \sqrt{(r_{\pm}^0)^2 + 4a}\,]\,/2 \quad . \tag{2.4}$$

Now, writing $h_\theta(a)$ in terms of r and β, we obtain

$$h_\theta(a) = K \int_0^{\beta_0} \int_{r_-}^{r_+} r^{p-1} \sin^{p-2}\beta\, f^*(r,\beta)\,dr\,d\beta \tag{2.5}$$

where $K = 2\pi \prod_{i=1}^{p-3} \int_0^{\pi} \sin^i t\, dt$ and

$$f^*(r,\beta) = \frac{1}{(\sqrt{2\pi})^p} e^{\{-(r^2-2r|\theta|\cos\beta + |\theta|^2)/2\}} \tag{2.6}$$

Using (2.5) one can establish the following lemma, which gives an expression for $\frac{\partial}{\partial a} h_\theta(a) = h_\theta'(a)$.

Lemma 1.

$$\frac{\partial}{\partial a} P_\theta(\theta \in C_\delta a) = K \int_0^{\beta_0} m(a,\theta,\beta)\,d\beta \tag{2.7}$$

where

$$m(a,\theta,\beta) = \sin^{p-2}\beta \left[\frac{r_+^p f^*(r_+,\beta)}{r_+^2 + a} - \frac{r_-^p f^*(r_-,\beta)}{r_-^2 + a}\right]$$

Proof. Differentiate (2.5) with respect to a. Note that β_0 is independent of a. Interchanging the order of differentiation and integration can be justified by the Bounded Convergence Theorem. Hence

$$h'(a) = K \int_0^{\beta_0} \frac{\partial}{\partial a} \int_{r_-}^{r_+} r^{p-1} f^*(r,\beta) \sin^{p-2}\beta\, dr\, d\beta$$

The fundamental theorem of calculus then gives $h_\theta'(a) = K \int_0^{\beta_0} m(a,\theta,\beta)d\beta$

where

$$(2.8)\quad m(a,\theta,\beta) = \sin^{p-2}\beta\ \{r_+^{p-1} f^*(r_+,\beta)(\tfrac{\partial}{\partial a})(r_+(a,\theta,\beta)) - r_-^{p-1}\ f^*(r_-,\beta)(\tfrac{\partial}{\partial a})(r_-(a,\theta,\beta))\}.$$

Consider first the case $|\theta| > c$. From (2.3), and the implicit function theorem, it follows that

$$(2.9)\quad (\tfrac{\partial}{\partial a})(r_\pm(a,\theta,\beta)) = [r_\pm(1+(a/r_\pm^2))]^{-1}$$

Substituting (2.9) into (2.8) establishes (2.7) for $|\theta| > c$. For $|\theta| \leq c$, (2.7) can be similarly established.

It is quite surprising that the derivative formula in Lemma 1 looks somewhat similar to (although much simpler than) the formula (2.21) of Hwang and Casella (1982), even though their formula is for the derivative of $P_\theta(\theta \in C_\delta a)$ with respect to $|\theta|$ rather than a. Consequently, the condition on a for domination in the main theorem below has a similar form.

Generalizations of (2.7) to other distributions are immediate. In fact if the density is $g(|x-\theta|^2)$, define $g^*(r,\beta) = g(r^2+|\theta|^2-2r|\theta| \cos \beta)$ and substitute g^* for f^* in (2.7). Lemma 1 then holds under a minor regularity condition on g (so that, in the proof, the order of differentiation and integration can be interchanged).

To show that $C_\delta a$ has higher coverage probability, our technique is to show that, for every θ, $h_\theta'(a) > 0$ for some a in an interval $(0,a^*]$, say. Since $h_\theta(0) = P(\theta \in C_X^0)$, this implies $P_\theta(\theta \in C_\delta) > P_\theta(\theta \in C_X)$ for every θ. In light of Lemma 1, $h_\theta'(a) > 0$ can be established if we can show that the integrand $m(a,\theta,\beta)$ is positive for almost every β. Note this is automatically satisfied for $|\theta| \leq c$ and every $a > 0$, since

$$m(a,\theta,\beta) = \sin^{p-2}\beta\ r_+^p\ f^*(r_+,\beta)/(r_+^2+a)$$

for $|\theta| \leq c$. In the proof of the following main theorems, we therefore only consider the case $|\theta| > c$.

Theorem 2.1.

The coverage probability of $C_{\delta}a$ is higher than the coverage probability of C_X^0 for every θ, provided $0 < a \leq a_0^*$. where $a_0^* > 0$ is the unique solution to

$$(2.10) \quad \left(\frac{c + \sqrt{c^2 + a_0^*}}{\sqrt{a_0^*}}\right)^{p-2} e^{-c\sqrt{a_0^*}} = 1 .$$

Proof. To show the positivity of the integrand m in (2.7) for $|\theta| > c$ and $0 < \beta < \beta_0$, one only needs to show

$$(2.11) \quad R \overset{\text{defn.}}{=} \frac{r_+^p f^*(r_+,\beta)}{r_-^p f^*(r_-,\beta)} \frac{r_-^2 + a}{r_+^2 + a} > 1 .$$

Therefore the theorem will be proved if one can establish (2.11) for $0 < \beta < \beta_0$ and $|\theta| > c$.

From (2.6), it follows that

$$(2.12) \quad R = \left(\frac{r_+}{r_-}\right)^p \exp\{-(r_+-r_-)(r_+-r_--2|\theta|\cos\beta)/2\} \frac{r_-^2+a}{r_+^2+a}$$

Equation (2.3) gives

$$(2.13) \quad 2|\theta|\cos\beta = r_+ u(r_+) + r_- u(r_-)$$

$$= r_+ + r_- - \frac{a}{r_+} - \frac{a}{r_-} .$$

By substituting (2.13) into (2.12) and by a little algebra, it can be easily shown that

$$(2.14) \quad R = s_{p-2}(t) \frac{1 + a/r_-^2}{1 + a/r_+^2}$$

where

$$(2.15) \quad s_m(t) \overset{\text{defn.}}{=} t^m \exp\{-a(t-t^{-1})/2\}$$

and

$$t \overset{\text{defn.}}{=} r_+/r_- \geq 1 .$$

For $|\theta| > c$ and $0 < \beta < \beta_0$, $[(1+ar_-^{-2})/(1+ar_+^{-2})] > 1$. Hence this, together with (2.14), will imply (2.11) provided one can establish $s_{p-2}(t) \geq 1$. Clearly $s_{p-2}(1) = 1$ and for $t \geq 1$, $s_{p-2}(t)$ either decreases or increases to a unique maximum and then decreases to zero. In either cases, to prove that $s_{p-2}(t) \geq 1$, it thus suffices to show $s_{p-2}(t^*) \geq 1$ where

(2.16) $$t^* = \max_{\substack{|\theta| > c \\ 0 \leq \beta \leq \beta_0}} r_+(a,\theta,\beta)/r_-(a,\theta,\beta).$$

For fixed $|\theta|$, r_+ is decreasing in β and r_- is increasing in β. Consequently t is decreasing in β, which implies

$$\sup_{0\leq\beta\leq\beta_0} t = \frac{r_+}{r_-}\Big|_{\beta = 0}$$

$$= \frac{|\theta| + c + \{(|\theta|+c)^2+4a\}^{1/2}}{|\theta| - c + \{(|\theta|+c)^2+4a\}^{1/2}} .$$

Direct differentiation shows that $\frac{d}{d|\theta|} \sup_{\beta\leq\beta_0} t$ equals

$$\frac{(|\theta|-c)(\ell^{1/2} - \ell^{-1/2}) + 2c(\ell^{1/2} - 1)}{[|\theta|-c+\{(|\theta|-c)^2+4a\}^{1/2}]^2}$$

where $\ell = \{(|\theta|-c)^2+4a\}/\{(|\theta|+c)^2+4a\} < 1$. For $|\theta| > c$, $\frac{d}{d|\theta|} \sup_{\beta<\beta_0} t < 0$ and consequently $t^* = (c+\sqrt{c^2+a})/\sqrt{a}$. Straightforward calculation shows $t^* - (t^*)^{-1} = 2c/\sqrt{a}$. Hence (2.11) will hold provided

(2.17) $$s^*(a) \overset{\text{defn.}}{=} \left(\frac{c+\sqrt{c^2+a}}{\sqrt{a}}\right)^{p-2} e^{-c\sqrt{a}} \geq 1 .$$

The function $s^*(a)$ is strictly decreasing in a. If one defines a_0^* to be as in (2.10), (2.11) hold for all a, $0 < a \leq a_0^*$. This proves the theorem.

The condition in (2.17) is very similar to the condition in (3.7) of Hwang and Casella (1982), except that the power $p-2$ in (2.10) is replaced by $(p-3)$

in Hwang and Casella (1982). The increase in power clearly gives us extra leverage and, consequently, the solution a_0^* to (2.10) is larger and our theorem is stronger, also covering the case $p=3$. In proving Theorem 2.1, we bounded the term $[1+ar_-^{-2}]/[1+ar_+^{-2}]$ in (2.14) by 1. If we deal with this term carefully, we can get an even stronger result as provided below.

Theorem 2.2.

The coverage probability of $C_\delta a$ is higher than the coverage probability of C_X^0 for every θ provided $0 < a \leq a^*$ where $a^* = \text{Min}\,(a_1^*, a_2^*)$, and $a_1^* > 0$ and $a_2^* > 0$ are solutions uniquely determined by

$$(2.18) \qquad \left(\frac{c/2 + \sqrt{(c/2)^2 + a_1^*}}{\sqrt{a_1^*}}\right)^{p-2} e^{-c\sqrt{a_1^*}/2} = 1$$

and

$$(2.19) \qquad \left(\frac{c + \sqrt{c^2+a_2^*}}{\sqrt{a_2^*}}\right)^{p-1} e^{-c\sqrt{a_2^*}} \left(\frac{\sqrt{c^2+4a_2^*} - c}{2\sqrt{a_2^*}}\right) = 1.$$

Proof. Again it is sufficient to show R, in (2.14), is greater than one for almost all θ and β. Write

$$R = s_p(t)\,(r_-^2+a)\,/\,(r_+^2+a)$$

where $s_p(t)$ is defined in (2.15). Using (2.4), straightforward calculation shows

$$\frac{r_-^2+a}{r_+^2+a} = \frac{r_-\sqrt{(r_-^0)^2+4a}}{r_+\sqrt{(r_+^0)^2+4a}},$$

which implies

$$(2.20) \qquad R = s_{p-1}(t)\,\sqrt{(r_-^0)^2+4a}\,/\,\sqrt{(r_+^0)^2+4a}\,.$$

Let $t'_{p-2} > 1$ be the unique solution to $s_{p-2}(t) = 1$. If the unique solution for $s_{p-2}(t) = 1$, $t \geq 1$, is one, define $t'_{p-2} = 1$. If $t \leq t'_{p-2}$, it follows from (2.14) that $R > 1$ for $|\theta| > c$ and $0 < \beta < \beta_0$ and the theorem is established.

If $t > t'_{p-2}$, we apply (2.20). It can be shown that the unique minimum of $\sqrt{(r_-^0)^2+4a} / \sqrt{(r_+^0)^2+4a}$ occurs at $\beta = 0$ and $|\theta| = (c^2+4a)^{1/2}$ which implies that

$$\sqrt{(r_-^0)^2 + 4a} / \sqrt{(r_+^0) + 4a} > [\sqrt{c^2 + 4a} - c] / (2\sqrt{a})$$

for all $|\theta| \neq (c^2+4a)^{1/2}$ and $\beta \neq 0$. This, together with (2.20), gives

$$R > s_{p-1}(t) \frac{\sqrt{c^2+4a} - c}{2\sqrt{a}} \tag{2.21}$$

for almost all θ and β. Since s_{p-1} either decreases, or increases to a unique maximum and decreases to 0, $R > 1$ will follow from (2.21) provided at the endpoints t'_{p-2} and t^*,

$$s_{p-1}(t)\Big|_{t=t'_{p-2}} \geq 2\sqrt{a} / (\sqrt{c^2+4a} - c) \tag{2.22}$$

and

$$s_{p-1}(t)|_{t=t^*} \geq 2\sqrt{a} / (\sqrt{c^2+4a} - c) . \tag{2.23}$$

(Recall the definition of t^* from (2.16).)

Since $s_{p-2}(t'_{p-2}) = 1$ and $s_{p-1}(t) = t\ s_{p-2}(t)$, (2.22) is equivalent to

$$t'_{p-2} \geq 2\sqrt{a} / (\sqrt{c^2+4a} - c) . \tag{2.24}$$

Recognizing $2\sqrt{a} / (\sqrt{c^2+4a} - c) \geq 1$, (2.24) is thus equivalent to

$$s_{p-2}(2\sqrt{a} / (\sqrt{c^2+4a} - c)) \geq 1 . \tag{2.25}$$

By direct substitution, (2.25) holds if and only if

$$\left(\frac{c/2 + \sqrt{(c/2)^2 + a}}{\sqrt{a}}\right)^{p-2} e^{-c\sqrt{a}/2} \geqq 1 . \tag{2.26}$$

The left hand side of (2.26) decreases in a . Hence if we let a_1^* be as defined by (2.18), (2.26) holds if and only if $a \le a_1^*$.

By a calculation similar to that which led to (2.17), it can be shown that (2.23) is equivalent to

$$(2.27)\qquad \left(\frac{c+\sqrt{c^2+a}}{\sqrt{a}}\right)^{p-1} e^{-c\sqrt{a}} \left(\frac{\sqrt{c^2+4a}-c}{2\sqrt{a}}\right) \ge 1 .$$

The left hand side of (2.27) is again decreasing in a, since $(c+\sqrt{c^2+a})(\sqrt{c^2+4a}-c)/2a$ is. Hence (2.27) holds if and only if $a \le a_2^*$. These establish the theorem.

Using a programmable calculator, the values of a^* can be obtained and are reported in Tables 1 and 2. For comparison purpose the upper bound a_0 in Hwang and Casella (1982) are also given in the same tables. (Recall the definition of a_0 from Section I.) For $p = 3$, our theorems provide domination results, while no result was given in Hwang and Casella (1982). In Tables 1 and 2, except when $p = 3$, $a^* = a_2^*$ and hence equation (2.19) is usually the more crucial equation than (2.18) in determining a^*. It can be shown analytically, that $a_2^* > a_0^*$, hence Theorem 2.2 usually provides larger bound than Theorem 2.1.

The increase in coverage probability of $C_\delta a$ with $a = a^*$ over that of $C_\delta a$ with $a = a_0$ is not expected to be large. This is due to the fact that $C_\delta a$ with $a = a_0$ has similar coverage probability as $C_\delta a$ with $a = p-2$. (See the numerical comparison in Hwang and Casella (1982).) It is expected that the coverage probability of $C_\delta a$ with $a = a^*$ is in between that of $a = a_0$ and $a = p-2$.

Theorem 2.2 could possibly be strengthened, if one could, instead of minimizing s_{p-1} and $[(r_-^0)^2+4a]^{1/2} / [(r_+^0)^2+4a]^{1/2}$ in (2.20) separately, find the minimum of R. Using the minimum, a better sufficient condition (say $0 < a \le a^{**}$) can be similarly established. (This upper bound a^{**} would be the largest bound that one could possibly obtain for the domination of $C_\delta a$ over C_X^0, if he would require the integrand in (2.7) to be everywhere nonnegative.) However, even if this can be done, the difference between a^{**} and a^* is minimal. We did not investigate a^{**}, because we do not have a consise formula for the minimum of R. However, we derive another upper bound a^{***} by setting $|\theta| = c$ and $\beta = 0$ and by solving $R = 1$. Clearly $a^* \leq a^{**} \leq a^{***}$. Using a programmable calculator, we calculate a^{***} and report the values in Tables 1 and 2. The difference between a^* and a^{***} is always less than 0.1. Therefore, for these cases, the increase of a^{**} over a^* cannot be greater than 0.1.

§3. Asymptotic expansion of the coverage probability and a necessary condition.

In this section, we establish an asymptotic expression (3.1) (as $|\theta| \to \infty$) of the coverage probability of $C_\delta a$.

Theorem 3.1. As $|\theta| \to \infty$,

$$(3.1)\qquad P(\theta \in C_\delta a) = (1-\alpha) - \frac{a}{2|\theta|^2}[1-\alpha-h(\alpha)][a-2(p-2)] + O(|\theta|^{-3}),$$

where $1-\alpha = P(|X-\theta|^2 \leq c)$,

$$(3.2)\qquad h(\alpha) = 1 - \alpha - c^p e^{-c^2/2} / [2^{\frac{p-2}{2}} \Gamma(\tfrac{p}{2})],$$

and $O(|\theta|^k)$ denotes a function bounded by a constant multiple of $|\theta|^k$. Applying Theorem 3.1, the following Corollary can be established easily.

Corollary 3.1.

If $a > 2(p-2)$, then $C_\delta a$ does not dominate C_X^0. In particular, $C_\delta a$ has smaller coverage probability for large $|\theta|$. If $a < 2(p-2)$, then $C_\delta a$ has higher coverage probability than C_X^0 for large $|\theta|$.

According to Corollary 3.1, if $C_\delta a$ dominates C_X^0, then $a \leq 2(p-2)$. It is also true that if $C_\delta a$ dominates C_X^0 then $a > 0$. Since if $a = 0$, $C_\delta a$ is the same as C_X^0 and $C_\delta a$ does not dominate C_X^0. Further if $a < 0$, it is straightforward to show that, for $|\theta| \leq c$, the event $\{\theta \in C_\delta a\}$ is properly contained in the event $\{\theta \in C_X^0\}$. This implies that the coverage probability of $C_\delta a$ is less than that of C_X^0 for $|\theta| < c$. We hence have the following necessary condition.

Corollary 3.2. A necessary condition for the domination of $C_\delta a$ over C_X^0 is $0 < a \leq 2(p-2)$.

The necessary condition for domination is not sufficient as suggested by our numerical study. However, the numerical study also indicates that the largest bound a_B, for the domination of $C_\delta a$ over C_X^0 for all $0 < a \leq a_B$, is close to $2(p-2)$. For $\alpha = 0.1$, $p = 3,15,17,19$ and 21 a_B is close to 2, 25,28,32, and 36, respectively.

Proof of Theorem 3.1. The following proof is similar, in some respects, to that of Theorem 3.3.1 in Berger (1980). Let

$$S_\theta = \{X: |\theta - \delta^a(X)| \leq c\} .$$

Hence

$$P_\theta(\theta \varepsilon C_\delta a) = P_\theta(X \varepsilon S_\theta) = \int_{S_\theta} \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}|X-\theta|^2} dX . \tag{3.3}$$

The scheme is to expand S_θ and $|X-\theta|^2$ in terms of some manageable forms and integrate out the dominating terms exactly. In the following expansion, keep in mind that $X \varepsilon S_\theta$ implies $|X-\theta|$ is bounded by some constant for all $|\theta|$.

First we deal with S_θ. Since we are considering only large $|\theta|$ we assume $|\theta| > c$. Consequently, $X \varepsilon S_\theta$ implies $|X|^2 > a$ and

$$\delta^a(X) = \left(1 - \frac{a}{|X|^2}\right) X ,$$

i.e., "+" in (1.1) can be omitted. Using the fact that

$$\frac{1}{|X|^2} = \frac{1}{|\theta|^2} - \frac{2\theta'(X-\theta)}{|\theta|^4} + O(|\theta|^{-4}) ,$$

we get

$$\theta - \delta^a(X) = \left[\theta - X + \frac{a\theta}{|\theta|^2} - \frac{a(\theta-X)}{|\theta|^2} + \frac{2a\theta\theta'(\theta-X)}{|\theta|^4}\right] + \underset{\sim}{O}(|\theta|^{-3}) \tag{3.4}$$

$$\underset{=}{\text{def}}'\text{n.} \quad Y + \underset{\sim}{O}(|\theta|^{-3}),$$

where $\underset{\sim}{O}(|\theta|^k)$ denotes some vector (or in general a matrix) with the property that the sum of the absolute values of all the elements is $O(|\theta|^k)$.

Equation (3.4) therefore implies

$$(3.5) \qquad |\theta - \delta^a(X)| = |Y|^2 + O(|\theta|^{-3}) .$$

Now consider $\exp(-|X-\theta|^2/2)$ in the integrand. Let I be the identity matrix and write

$$Y = [(1 - \frac{a}{|\theta|^2}) I + \frac{2a\theta\theta'}{|\theta|^4}] (\theta - X) + \frac{a\theta}{|\theta|^2} .$$

Note

$$(3.6) \qquad (aI + bww')^{-1} = \frac{1}{a}(I - \frac{b}{a+bw'w} ww')$$

$$(3.7) \qquad \det(aI + bww') = a^p(1 + \frac{b}{a} w'w)$$

for any constants a and b and any column vector w as long as the expressions make sense. Equation (3.6) gives

$$[(1 - \frac{a}{|\theta|^2})I + \frac{2\,a\,\theta\theta'}{|\theta|^4}]^{-1} = (1 + \frac{a}{|\theta|^2})I - \frac{2a}{|\theta|^4}\theta\theta' + \underset{\sim}{O}(|\theta|^{-4})$$

and

$$(3.8) \quad \theta - X = [(1 + \frac{a}{|\theta|^2}) I - \frac{2a}{|\theta|^4}\theta\theta'] (Y - \frac{a\theta}{|\theta|^2}) + \underset{\sim}{O}(|\theta|^{-4}),$$

which, in turns, give

$$|\theta - X|^2 = |Y|^2 - \frac{2a\theta'Y}{|\theta|^2} + \frac{2a|Y|^2}{|\theta|^2} + \frac{a^2}{|\theta|^2} - \frac{4a(\theta'Y)^2}{|\theta|^4} + O(|\theta|^{-3}) ,$$

and

$$(3.9) \quad e^{-\frac{1}{2}|\theta - X|^2} = e^{-|Y|^2/2}\{1 + \frac{a\theta}{|\theta|^2} - \frac{a|Y|^2}{|\theta|^2} - \frac{a^2/2}{|\theta|^2} + \frac{2a(\theta'Y)^2}{|\theta|^4} + \frac{a^2(\theta'Y)^2}{2|\theta|^4}$$

$$+ O(|\theta|^{-3}) \}$$

Let $f(\cdot)$ be

$$f(y) = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}|y|^2}$$

Using (3.5) and (3.9), and changing the variable X into Y in (3.3), we get

(3.10) $$P_\theta(\theta \varepsilon C_\delta a)$$

$$= \int_{|Y|^2 \leq c^2 + O(|\theta|^{-3})} f(Y)[1 + \frac{a\theta'Y}{|\theta|^2} - \frac{a|Y|^2}{2|\theta|^2} + \frac{2a(\theta'Y)^2}{|\theta|^4} + \frac{a^2(\theta'Y)^2}{2|\theta|^4} + O(|\theta|^{-3})]Jdy$$

where, by (3.7) and (3.8), the Jacobian is

$$J = |\det[(1 + \frac{a}{|\theta|^2})I - \frac{2a}{|\theta|^4}\theta\theta']| + O(|\theta|^{-4})$$

$$= 1 + \frac{a(p-2)}{|\theta|^2} + O(|\theta|^{-4}) .$$

This, together with (3.10), implies

$$P_\theta(\theta \varepsilon C_\delta a) = \int_{|Y| \leq c} f(Y)[1 + \frac{a\theta'Y}{|\theta|^2} - \frac{a|Y|^2}{2|\theta|^2} + \frac{2a(\theta'Y)^2}{|\theta|^4}$$

$$+ \frac{a^2(\theta'Y)^2}{2|\theta|^4} + \frac{a(p-2)}{|\theta|^2}]dY + O(|\theta|^{-3}) .$$

Let $h(\alpha)$ be as in (3.2). Straightforward calculation shows

$$\int_{|Y| \leq c} |Y|^2 f(Y)dY = ph(\alpha)$$

$$\int_{|Y| \leq c} (\theta'Y)f(Y)dY = 0$$

and

$$\int_{|Y| \leq c} (\theta'Y)^2 f(Y)dY = |\theta|^2 h(\alpha).$$

Therefore the coverage probability equals

$$1 - \alpha - \frac{aph(\alpha)}{|\theta|^2} - \frac{a^2}{2}\frac{1-\alpha}{|\theta|^2} + \frac{a^2}{2}\frac{h(\alpha)}{|\theta|^2} + \frac{2ah(\alpha)}{|\theta|^2} + \frac{a(p-2)(1-\alpha)}{|\theta|^2} + O(|\theta|^{-3}) .$$

This establishes Theorem 3.1.

REFERENCES

Berger, J.O. (1980), "A Robust Generalized Bayes Estimate and Confidence Region for a Multivariate Normal Mean," Annals of Statistics 8, 716-761.

Brown, L.D. (1966), "On the Admissibility of Invariant Estimators of One or More Location Parameters," Annals of Mathematical Statistics, 37, 1087-1136.

Casella, G. (1980), "Coverage Probabilities for Multivariate Normal Confidence Sets Centered at James-Stein Estimators," Technical Report, Biometrics Unit, Cornell University.

Casella, G. and Hwang, J.T. (1982), "Empirical Bayes Confidence Sets for the Mean of a Multivariate Normal Distribution". Technical Report, Biometrics Unit, Cornell University.

Efron, B. and Morris, C. (1973), "Stein's Estimation Rule and its Competitors - an Emperical Bayes Approach," Journal of the American Statistical Association, 68, 117-130.

Faith, R.E. (1976), "Minimax Bayes Set and Point Estimators of a Multivariate Normal Mean, "Technical Report #66, University of Michigan.

Hwang, J.T., and Casella, G. (1982), "Minimax Confidence Sets for the Mean of a Multivariate Normal Distribution," Technical Report, Cornell University. To appear in Annals of Statistics, September 1982.

James, W. and Stein, C. (1961), "Estimation with Quadratic Loss," Proceedings of the Fourth Berkeley Symposium, University of California Press, 1, 361-379.

Joshi, V.M. (1967), "Inadmissibility of the Usual Confidence Sets for the Mean of a Multivariate Normal Population," Annals of Mathematical Statistics, 38, 1868-1875.

Joshi, V.M. (1969), "Admissibility of the Usual Confidence Sets for the Mean of a Univariate or Bivariate Normal Population," Annals of Mathematical Statistics, 40, 1042-1067.

Stein, C. (1956), "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", Proc. Third Berkeley Sym. Math. Statist. Prob.1, 197-206. Univ. of California Press.

Table 1: Upper bounds on the range of a, α = .10.

	a_0	a_0^*	a^*	a^{***}
3	0	.580	.815	.892
4	.538	1.339	1.668	1.710
5	1.275	2.132	2.483	2.536
6	2.058	2.942	3.306	3.366
7	2.860	3.760	4.133	4.199
8	3.674	4.585	4.964	5.035
9	4.495	5.413	5.797	5.872
10	5.321	5.245	6.633	6.711
11	6.151	7.079	7.470	7.551
12	6.984	7.915	8.310	8.393
13	7.819	8.754	9.150	9.235
14	8.656	9.593	9.992	10.079
15	9.495	10.434	10.834	10.923
16	10.335	11.276	11.678	11.768
17	11.177	12.119	12.522	12.614
18	12.019	12.963	13.368	13.461
19	12.863	13.808	14.214	14.308
20	13.707	14.653	15.060	15.155
21	14.552	15.500	15.907	16.003
22	15.398	16.346	16.755	16.852
23	16.244	17.194	17.603	17.701
24	17.092	18.042	18.452	18.550
25	17.939	18.890	19.301	19.400

Table 2. Upper bounds on the range of a, $\alpha = .05$.

	a_0	a_0^*	a^*	a^{***}
3	0	.537	.784	.812
4	.500	1.265	1.570	1.602
5	1.207	2.036	2.367	2.408
6	1.966	2.828	3.173	3.223
7	2.750	3.631	3.987	4.042
8	3.548	4.441	4.806	4.866
9	4.354	5.258	5.628	5.692
10	5.168	6.078	6.454	6.521
11	5.987	6.902	7.282	7.352
12	6.809	7.729	8.112	8.185
13	7.634	8.558	8.944	9.019
14	8.462	9.389	9.777	9.854
15	9.292	10.222	10.612	10.691
16	10.124	11.056	11.448	11.529
17	10.957	11.891	12.285	12.368
18	11.792	12.728	13.124	13.207
19	12.629	13.566	13.963	14.048
20	13.466	14.405	14.803	14.889
21	14.304	15.244	15.643	15.731
22	15.143	16.084	16.485	16.573
23	15.983	16.926	17.327	17.417
24	16.824	17.767	18.170	18.260
25	17.665	18.610	19.013	19.105